Matter Over Mind? How the Acceptance of Digital Entities Depends on Their Appearance, Mental Prowess, and the Interaction Between Both

Jan-Philipp Stein[1], Markus Appel[1], Alexandra Jost[2], & Peter Ohler[2]

[1] Psychology of Communication and New Media, Julius Maximilian University of Würzburg

[2] Chair of Media Psychology, Chemnitz University of Technology

Abstract

Digital technologies are advancing rapidly, growing to be more human-like and intelligent by the day. However, research shows that a machine's resemblance to humans can reach a critical level, which makes it seem uncanny to observers. While scholars have discussed this effect in terms of both human-like appearances and mental abilities, a potential interaction between the two aspects has hardly been addressed in literature. We designed a two-factorial experiment to overcome the identified research gap, introducing participants to digital agents with varying embodiment (text interface/human rendering) and mental capacity (simple algorithms/complex artificial intelligence). Our results show that the interaction of both factors indeed affects participants' experience in a crucial way: Whereas an agent based on simple algorithms only evokes discomfort when embedded in a human-like body, the artificial intelligence is always perceived as eerie, regardless of its embodiment. Yet, additional findings raise doubts on the unidimensionality of participants' affective response.

Keywords: *digital agent, artificial intelligence, aversion, uncanny valley, mind attribution, embodied technology*

Matter Over Mind? How the Acceptance of Digital Entities Depends on Their

Appearance, Mental Prowess, and the Interaction Between Both

Even though some of the ways past authors envisioned the 21st century are likely to remain obscure fantasies, real life does catch up with science fiction from time to time, not least regarding the possibilities of human–computer interaction. From virtual agents encouraging customers to buy a product to companion robots giving solace to the elderly, people are getting used to more and more technologies that can be addressed like a regular person—similar to how Luke Skywalker talked to his android C-3PO in the popular "Star Wars" movies.

While the tendency to humanize the non-human (a process called *anthropomorphization*) is actually much older than modern age technology, computers make it particularly easy for the human brain to anthropomorphize them. Interpreting the responsiveness of digital machinery as social affordance, people are hardwired to fall back on the same heuristics they usually apply to human-to-human contexts (Nass & Moon, 2000). Considering that new digital technologies make it increasingly easy to perceive them as complex interaction partners, this so-called *computers are social actors* paradigm (CASA; Nass, Steuer, & Tauber, 1994; Reeves & Nass, 1996) has only been growing in relevance, and will likely continue to do so for the foreseeable future.

Apart from the technical advancement itself, however, it remains less certain if reality will also mirror fiction in terms of the societal acceptance of new digital inventions. After all, popular culture has made a habit of distorting technological progress into the herald of disaster (Dinello, 2005; Ferrari, Jetten, & Paladino, 2016), especially in the world of Western fiction. Based on its numerous stories about revolting man-made beings—from disobedient broomsticks to malevolent spaceship computers—cultural scientists have even attested Western culture a so-called "Frankenstein Syndrome" (Kaplan, 2004): a deep-rooted fear of self-controlled technology. While the premise itself may sound vaguely prosaic,

psychological research actually corroborates it on an empirical level, as European and North American study participants tend to express limited trust towards autonomous machines (e.g., Bartneck et al., 2007; Gnambs & Appel, 2019). Even more explicitly, recent experiments have revealed that people's aversion to advanced technology is directly predicted by their engagement with dystopian science fiction (Sundar, Waddell, & Jung, 2016; Young & Carpenter, 2018)—an observation in line with the common notion of media-transmitted experience as a determinant of real-life attitudes (Potter, 2013).

For contemporary HCI developers and researchers, the potential fear of autonomous technology has arguably become one of the most pressing issues in their line of work. Despite the topic's new-found relevance, however, literature shows that efforts to understand *technophobia* (i.e., the fear of complex devices) are actually much older, reaching back far into previous decades. A particularly prominent remnant in this regard is the *uncanny valley* hypothesis, which was developed by Japanese roboticist Masahiro Mori nearly fifty years ago (Mori, 1970; Mori, MacDorman, & Kangeki, 2012). Having observed people's fearful reactions to artificial limbs, Mori suggested that for a human-like replica to be well-accepted, increasing its realism would only be beneficial up to a certain threshold, beyond which observers would inevitably experience a cold and eerie sensation. Based on this assumption, Mori further argued that certain types of puppets, machines, and robots inhabited the same mental space as corpses or zombies, with their ambiguous nature as half-dead/half-alive entities prompting innate defense mechanisms. Although it has to be noted that the author never conceived of his hypothesis as a scientific statement (Robertson, 2018), Mori's impressionistic approach to technical innovation has still provided scholars with a comprehensible and widely acclaimed framework for many years.

New Shapes of the Uncanny

In its original form, the uncanny valley strictly revolves around physical factors, which may explain why most of the subsequent research on the phenomenon also focused on

design features, movement patterns, and the general aesthetics of human-like machines (e.g., Hanson, 2005; Kwak, Kim, & Choi, 2017; Rosenthal-von der Pütten & Krämer, 2014). In recent years, however, academic interest has started to shift towards a new facet of human-like artificiality, namely the intelligence and mental abilities of digital systems (Gray & Wegner, 2012; Stein & Ohler, 2017). Considering that contemporary digital innovations continue to leave the restrictions of human programming behind—for instance by employing deep learning procedures to gain their own understanding of the world (LeCun, Bengio, & Hinton, 2015)—they also seem to approach a limit of desirable human likeness similar to the one predicted in Mori's initial hypothesis. Indeed, whereas simpler machines with a partial similarity to humans are usually seen as charming or enjoyable (Khan, 1998; Lee et al., 2006), a pioneering study by Gray and Wegner (2012) showed that more advanced digital minds might be met with strong apprehension. In another recent experiment, Stein and Ohler (2017) found that an artificial intelligence with the ability to simulate human empathy fell into an "uncanny valley of mind," being perceived as eerie and highly unnatural by participants. As a possible explanation for this result, the authors proposed that their stimuli had evoked a threat to people's concept of *human uniqueness*, which often revolves around humanity as the crown of all creation (e.g., Gee, Browne, & Kawamura, 2005; Schultz, Zelezny & Dalrymple, 2000). In fact, cultural observations suggest that the Christian understanding of a uniquely human nature still informs the worldview of many inhabitants of Western industrial nations (Kaplan, 2004), even despite the ongoing decline of organized religion (Altemeyer, 2009). While the conceptualization of this distinctly human essence may vary with each person's specific philosophical stance, research has shown that most people consider aspects such as higher-order cognition (Leyens et al., 2001), emotional experience (Waytz et al., 2010), emotional warmth (Gray & Wegner, 2012), and sophisticated processing of social cues (Pagel, 2012; Vogeley & Bente, 2010) as explicitly human traits. In all probability, the validity of these criteria will keep experts from numerous research areas (including neuroscience, psychology,

and zoology) occupied for the near future; nevertheless, they still provide a veritable basis for people's current acceptance of sophisticated technology. Accordingly, empirical studies have shown that machines with the abovementioned abilities are typically perceived as a symbolic threat to human identity and therefore disliked by observers (Jetten et al., 1995; Yogeeswaran et al., 2016; Zlotowski, Yogeeswaran, & Bartneck, 2017). Even more so, recent research suggests that people may begin to fear for their physical safety or worry about losing jobs and resources once human-like machinery begins to display autonomous and emotional behavior (Waytz & Norton, 2014). Considering this connection between a technology's level of sophistication and people's aversion to it, it comes as no surprise that participants tasked with bullying or destroying complex robots have shown little restraint in the according experiments (Bartneck et al., 2007; Keijsers & Bartneck, 2018).

Interestingly, recent evidence suggests that the described effects might turn out quite differently once a digital mind gets concealed within an actual human body. Proposing the so-called *echoborg condition*, Corti and Gillepsie (2015) asked participants to interact with human confederates, who were secretly receiving their answers from artificial chatbots. By these means, the authors observed that most participants failed to detect the presence of an artificial mind, especially when compared to a group that used a text interface to talk to the supposedly human interlocutor. Arguably, this result suggests that the nature of a digital agent's embodiment may fundamentally alter the way its mind is perceived—a sentiment echoed by similar, recent studies comparing text-based chatbots to human-like virtual bodies (Ciechanowski et al., 2018).

The Current Study

People's psychological resistance against machines with a highly human-like exterior has been replicated by numerous studies, and more recently, research has started to uncover similar reactions towards those with human-like "interior" (i.e., cognitive and affective abilities) as well. A potential interplay of both findings, however, has hardly been tackled by

previous scientific efforts. While a couple of studies have demonstrated that artificial minds with anthropomorphic bodies are perceived as more intelligent and emotionally capable (Broadbent et al., 2013; Ferrari, Paladino, & Jetten, 2016; Krach et al., 2008; Powers & Kiesler, 2006), or that human "echoborgs" might mitigate the artificiality of digital agents (Corti & Gillepsie, 2015), we do not know of any work that experimentally connects these interaction effects to actual measures of technology acceptance. To overcome this research gap, we conducted a two-factorial experiment, investigating how a digital agent's appearance and mental abilities, as well as the potential interaction between both factors might influence observers' experience of eerie, unpleasant feelings.

In accordance with the abovementioned findings on enhanced mind perceptions in anthropomorphic entities, we first hypothesized:

H1:   Presenting a digital agent in the form of a human-like rendering instead of a bodiless text interface will lead to stronger attributions of mind.

Secondly, having reviewed recent literature on the "uncanny valley of mind," which proposes a strong aversion towards complex supercomputers (Gray & Wegner, 2012) and emotionally capable artificial intelligence (Stein & Ohler, 2017; Appel et al., 2020), we further assumed:

H2:   A digital agent that is characterized as a complex artificial intelligence will be perceived as more aversive than an agent relying on simple algorithms.

Lastly, we expected that the relationships postulated in H1 and H2 would inevitably reinforce each other, leading to an especially strong culmination of aversive feelings:

H3:   A digital agent that possesses a human-like appearance and is characterized as a complex artificial intelligence will be perceived as particularly aversive.

On an exploratory side note, we decided to also investigate a more pragmatic aspect of technology use, namely the utility ascribed to the digital agent. While it seemed logical that a more complex system, both in terms of embodiment and mental prowess, would be perceived as more useful, we also had to consider the possibility that the hypothesized aversion against artificial human likeness would manifest in the form of reduced utility ratings. Due to these conflicting assumptions, we decided to pose a research question instead of a directional hypothesis:

RQ: How does a digital agent's embodiment, mental prowess, and the interaction between both influence its perceived utility?

Method

We used the web tool *SoSciSurvey* to create an online experiment with four between-subject conditions. More specifically, we presented participants with (**a**) one of two self-created information texts, describing either a very simple conversational agent or a refined artificial intelligence, followed by (**b**) one of two video conditions, portraying the introduced system as either a bodiless text interface or a refined human-like rendering. Figure 1 depicts the resulting 2×2 between-subject design. SoSciSurvey's randomization feature was used to assign participants to one of the four experimental conditions.

Participants

An a priori calculation of minimum sample size—using G*Power software (Faul et al., 2003) and assuming 80% power for effects of at least $f = 0.25$—suggested a minimum number of 128 participants. With the help of local social media groups and mailing lists, we recruited 214 German-speaking participants (78.5% female; age $M = 23.85$ years, $SD = 6.77$) for the current study. Unfortunately, initial data cleansing showed that 80 participants had not watched the provided stimulus video completely or interrupted their filling out of our questionnaire for a longer amount of time, so that we had to exclude them from further

analyses. Consequently, the final sample consisted of 134 participants (84.3% female; age

$M = 23.17$ years, $SD = 5.91$), which we still deemed sufficient for our purposes according to

the conducted sample size calculation. Looking into the demographic distribution of our final

sample, we found most of the participants to be enrolled as university students (89.6%) or

currently employed (9.0%).

Complying with academic standards, each participant gave informed consent at the

beginning of our online questionnaire, having received comprehensive information about the

voluntariness and anonymity of his or her participation. In order to thank them for their time

and effort, all participants received a ticket for a gift raffle of two €25 Amazon vouchers.

Stimulus Materials

According to the developed 2×2 design, we required stimuli for four distinct

experimental conditions. To keep full control over the manipulation, we decided to compose

the necessary materials ourselves, using existing resources about current technological

possibilities as well as our own fictional pieces of information.

For the manipulation of *portrayed agent mind*, we created two written vignettes with a

length of approximately 300 words, allegedly describing a "currently widely acclaimed form

of conversational agent." The first version (*simple algorithm* condition) introduced an agent

with an extremely limited behavioural range, which "fully depended on rigorous

programming prior to its use." The text further explained that the agent, while capable to give

"more than one thousand answers," consisted of little more than basic question–response

algorithms, far removed from actual autonomy or emotionality. In contrast to this, the second

text (*complex artificial intelligence* condition) described the digital agent as a "marvel of

current neural network technology." In this vignette, the artificial system was framed as a

"surprisingly empathic and emotional interaction partner," whose mind resembled the human

brain close enough to "truly grasp what users mean—all by itself." Moreover, the text

emphasized that the agent's mental abilities were constantly evolving due to its embedded

deep learning procedures, "resulting in an artificial mind that might manifest new, unexpected character traits in the long run."

The manipulation of our second experimental factor, *portrayed agent embodiment*, was achieved in the form of different video clips presented directly after the abovementioned introduction texts. In the first condition (*text interface*), participants watched a brief video recording of an on-screen conversation with the chatbot "Cleverbot" (Carpenter, 2018)—a simple web interface (Figure 2) that allows talking to a digital agent via written messages. Lacking any visual features, Cleverbot merely provides its users with a plain text field, replying to each entered message with a short written answer. To increase the plausibility of our depiction, we prepared two slightly different edits of the Cleverbot conversation according to the agent mind texts that were presented beforehand. Whereas the Cleverbot video shown after the *simple algorithm* vignette depicted more limited, neutral answers (e.g., user: "Can you help me?" – Cleverbot: "Depends on what you need."), the edit shown after the *complex artificial intelligence* vignette showcased sophisticated conversational abilities, including emotional and humorous responses (user: "Do you feel sad?" – Cleverbot: "Sometimes, what about you?"; "Forever young, I want to be forever young.").

For our second embodiment condition (*human rendering*), on the other hand, we edited together excerpts from a YouTube demo reel of the digital agent "Zoe," a product of the New Zealand company SoulMachines (Figure 3). According to industry insiders (e.g., Griffin, 2019; Vlahos, 2019), SoulMachines' creations rank among the most realistic real-time renderings of human heads currently available, equipped with emotional facial expressions and authentic human-like voices. Again, we decided to edit Zoe's presentation in a subtle manner to match the preceding information texts. While the resulting edits did not differ in terms of Zoe's visual presentation, they either portrayed her as an agent with limited mental abilities (including verbal statements such as "If you don't like talking to me, you can ask to speak to a real person.") or as an emotional, sophisticated conversation partner (e.g., "I

like it when you're honest."). These slightly different videos were then presented according to the respective agent mind vignette each participant had received beforehand.

Measures

Following the presentation of our textual and visual stimuli, participants had to fill in various measures that served as dependent variables and control variables in our study.

Aversion. The operationalization of the outcome variable in Mori's uncanny valley theory has been the subject of controversial debate, being linked to numerous constructs such as unfamiliarity, eeriness, disgust, lack of warmth—or, in more general terms, *aversion*. To account for this conceptual ambiguity, we composed our assessment of participants' affective response from several well-established sources, striving for a multi-dimensional approach to the phenomenon in question.

In recent years, the combination of two measures developed by Ho and MacDorman (2017) has become a particularly prominent approach to measuring the uncanniness of technology, so that we decided to use them as a starting point for our aversion measurement. Ho and MacDorman's first scale, *eeriness*, consists of nine items (e.g., "dull–freaky," "predictable–eerie," "bland–uncanny"), which were presented as 7-point semantic differentials. We observed a very good internal consistency for the measure, Cronbach's α = .89. *Attractiveness*, the authors' second scale, comprises of four items (e.g., "repulsive–agreeable," "messy–sleek"), for which we also calculated a good internal consistency of Cronbach's α = .84. Despite the scale's name suggesting an emphasis on visual attributes, the authors note that they conceived of the measure as an affective dimension, proposing it as a valuable (and conceptually discriminant) addition to the eeriness scale. Complying with this recommendation, we added both scales to the current study.

Next, we decided to complement Ho and MacDorman's scales with an assessment of the *emotional warmth* perceived in the agent, which, in our opinion, was not sufficiently expressed by the first two aversion measures. In fact, Ho and MacDorman (2010) themselves

note that their two scales remain conceptually distinct from the general valence of social perceptions, which is typically described as "likability" or "warmth." However, since we were also interested in this aspect of human–machine interaction, we complemented our measurement by adapting four items from Reysen's Likability Scale (2005; e.g., "The digital agent is warm," "The digital agent is friendly"; 1 = *strongly disagree*, 7 = *strongly agree*). Not only did the resulting emotional warmth index achieve good internal consistency, Cronbach's α = .81, an exploratory factor analysis further indicated the homogeneity of our newly constructed measure, with a single factor accounting for 63.6% of the total variance. Last but not least, we completed our operationalization of participants' aversion with a focus on their behavioural intentions, which were assessed with two additional items from Reysen's Likability questionnaire. Asking participants about the extent to which they would "like to be co-workers with the agent" and "ask the agent for personal advice" (1 = *strongly disagree*, 7 = *strongly agree*), we assembled an *interest in future interactions* scale, observing a slightly less-than-ideal internal consistency of Cronbach's α = .69, but gathering clear evidence for the measure's homogeneity (with 76.5% of variance contained within a single factor).

Utility. The General Impressions of Humanoids questionnaire provided by Kamide and colleagues (2012) offers sub-scales for various aspects of human-robot-interaction. Based on our research propositions, we chose to adapt the three items of their *utility* sub-scale to our specific scenario, exchanging the word "robot" with "agent" (e.g., "I can't understand why this agent is necessary," "I can't find a purpose to use this agent"). The three items were presented on 7-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*), achieving near-excellent internal consistency, Cronbach's α = .88.

Perceived human likeness. To find out whether our self-created information materials had worked in manipulating participants' perceptions of human likeness in the digital agent, we constructed two indices, namely a *mental human likeness* scale (2 items; "How much human nature did you perceive in this digital mind?," "To which extent would

you ascribe real feelings to this agent?"), and a *visual human likeness* scale (2 items; "As how human-like did you consider the agent's appearance?," "As how human-like would you describe the agent's design?"). Both resulting indices showed acceptable internal consistency, Cronbach's α of .79 and .76, respectively. Exploratory factor analyses further indicated that both measures were clearly homogenous, with more than 80% of the total variance expressed in a single factor in each case.

Controlling for participants' technology expertise and technology concern.
According to previous research, people's evaluation of digital technology is heavily influenced by novelty effects, which might manifest in both positive and negative ways (e.g., Creed, Beale, & Cowan, 2015; Tinwell & Grimshaw, 2009). Acknowledging this circumstance, we decided to measure participants' previous experience with digital agents in order to include it as a covariate in our statistical analyses. Two items were self-composed for this purpose ("How would you rate your experience with digital agents?", "How often do you play video games?"; 1 = *very low/never*, 5 = *very high/very often*) and averaged into an *expertise with digital agents* scale. The resulting index proved to be of acceptable internal consistency, Cronbach's α = .75, whereas an exploratory factor analysis suggested the measure's homogeneity, showing that a single factor accounted for 80.5% of the variance in participants' answers.

Current technology acceptance literature postulates that people's cultural socialization can contribute to a macro-level attitude towards autonomous technology (Stein, Liebold, Ohler, 2019; Young & Carpenter, 2018)—or, in the context of human-like replicas, a specific *uncanny valley sensitivity* (MacDorman & Entezari, 2015)—which then modulates the acceptance of complex digital creations. In order to control for this attitudinal influence, we decided to measure people's *concern about autonomous technology* with a questionnaire dedicated to this purpose by Stein and colleagues (2019). Consisting of 13 items (e.g., "The idea that machines will one day have the same abilities as real humans will hopefully never

come true," "…does not change the value of humanity at all") rated on a 7-point scale, the measure achieved an excellent internal consistency, Cronbach's α = .91.

Results

For a descriptive overview of our obtained results, Table 1 shows the means and standard deviations for all scales, whereas Table 2 collects the zero-order correlations between the measured variables. Moreover, we would like to refer readers to Table 3 for a concise overview on the conducted hypothesis tests.

Human Likeness Perceptions

Focusing on hypothesis H1, we conducted a multivariate analysis of covariance (MANCOVA) with participants' perception of the agent's mental and visual human likeness as dependent variables, and the two experimental factors (portrayed agent mind and portrayed agent embodiment) as between-subject factors. Participants' previous experience with digital agents, as well as their general concern about autonomous technology were added as covariates. By inspecting histograms and scatterplots, calculating Mahalanobis' distance, and conducting a Box's M test, we made sure that all necessary assumptions for the procedure were met (Field, 2013).

The analysis resulted in a significant and very strong multivariate main effect of portrayed agent embodiment, $F(2,127) = 53.63$, $p < .001$, Wilk's $\Lambda = 0.54$, $\eta_p^2 = .46$. In contrast to this, neither the multivariate main effect of portrayed agent mind ($p = .81$), nor the interaction between portrayed agent mind and embodiment ($p = .21$) turned out to be significant.

Following up the significant multivariate result for portrayed agent embodiment with univariate analyses, we found that this experimental factor had actually affected both dependent variables in a significant way. First, participants' rating of the agent's visual human likeness was clearly affected by our visual presentation of different agent embodiments, $F(1,128) = 132.76$, $p < .001$, $\eta_p^2 = .39$. Unsurprisingly, participants who had

watched a video clip of the text interface Cleverbot indeed rated the system's appearance as much less human-like ($M = 3.32$, $SD = 1.39$) than those presented with the video of the human rendering Zoe ($M = 5.40$, $SD = 1.13$). Second, we observed that the portrayed agent embodiment had also exerted a significant influence on participants' attributions of mental human likeness, albeit to a much smaller extent, $F(1,128) = 7.45$, $p = .04$, $\eta_p^2 = .03$. Specifically, it was revealed that watching the text interface video had led to slightly lower mental human likeness ratings ($M = 3.51$, $SD = 1.36$) than being introduced to the human rendering ($M = 3.99$, $SD = 1.34$). As such, we give a (cautiously) positive answer to H1, noting that a digital agent with human-like visual features was indeed attributed with a more human-like mind—regardless of the cover story that was given regarding its actual mental prowess (see Table 3).

Aversion

As the next step of our data analysis, we conducted another MANCOVA with our obtained aversion measures (eeriness, attractiveness, emotional warmth, interest in future interactions) as dependent variables and both experimental factors (portrayed agent mind and portrayed agent embodiment) as between-group factors. Again, we added participants' previous experience with digital agents and their general level of concern about autonomous technology as covariates. Checking our data for their suitability concerning the planned procedure, we examined that all assumptions for MANCOVA were met.

The multivariate procedure yielded significant main effects for both portrayed agent mind, $F(4,125) = 2.55$, $p = .04$, Wilk's $\Lambda = 0.93$, $\eta_p^2 = .08$, and portrayed agent embodiment, $F(4,125) = 10.20$, $p < .001$, Wilk's $\Lambda = 0.75$, $\eta_p^2 = .25$. Additionally, a significant multivariate interaction effect between portrayed agent mind and embodiment could be observed, $F(5,124) = 4.36$, $p < .01$, Wilk's $\Lambda = 0.88$, $\eta_p^2 = .13$.

In our subsequent univariate analyses of covariance (ANCOVAs), we first explored the discovered main effect of portrayed agent mind. Focusing on the DVs attractiveness ($p =$

.17), emotional warmth ($p = .91$), and interest in future interactions ($p = .80$), no significant

findings emerged from our data. However, regarding the DV perceived eeriness, a moderately

strong influence of portrayed agent mind could be observed, $F(1,128) = 5.32$, $p = .01$, $\eta_p^2 =$

.05. Indeed, whereas the textual introduction to an advanced artificial intelligence resulted in

above-average feelings of eeriness ($M = 4.21$, $SD = 1.00$), the vignette about an agent relying

on simple algorithms resulted in an average eeriness rating below the scale's midpoint ($M =$

3.69, $SD = 0.98$). Considering our multi-faceted operationalization of aversion, however, we

only give a cautiously positive answer to H2. Since only one of the four aversion variables

was subject to a significant univariate effect, we ask that our finding is not generalized

carelessly.

Next, we used our subsequent ANCOVAs to disentangle the significant multivariate

effect of portrayed agent embodiment. While the investigation of the DVs emotional warmth

($p = .24$) and interest in future interactions ($p = .22$) remained without statistically noteworthy

results, we found that the two other ratings were significantly influenced by the agent's visual

presentation, both eeriness, $F(1,128) = 20.82$, $p < .001$, $\eta_p^2 = .17$, and attractiveness, $F(1,128)$

$= 16.66$, $p < .001$, $\eta_p^2 = .16$. Referring readers to our obtained group means as shown in Table

1, we report that perceiving the agent as a human-like rendering led to significantly higher

attractiveness scores, but also to a much more pronounced experience of eeriness than

viewing the video of a simple text interface.

Concluding our analysis of participants' aversion, we scrutinized the observed

multivariate interaction between portrayed agent mind and portrayed agent embodiment in

our univariate ANCOVAs. Doing so, we noticed that the interaction between both factors had

significantly affected the evaluation of eeriness, $F(1,128) = 4.47$, $p = .02$, $\eta_p^2 = .04$, and

emotional warmth, $F(1,128) = 7.09$, $p = .01$, $\eta_p^2 = .05$. As illustrated by Figure 4, participants

had experienced an artificially intelligent agent as rather eerie regardless of its embodiment

(*text interface* condition: $M = 3.99$, $SD = 1.08$; *human rendering* condition: $M = 4.41$, $SD =$

0.89)—while for agents based on simple algorithms, a human-like embodiment was seen as much eerier ($M = 4.36$, $SD = 0.73$) than a text interface ($M = 3.19$, $SD = 0.84$). Regarding the interaction effect on perceived emotional warmth, on the other hand, we observed an "inconsistency bias" (Figure 5): Mismatches of mental and visual complexity had resulted in slightly lower ratings, both for a simple algorithms agent in a human-like body ($M = 4.40$, $SD = 1.00$) and for an artificial intelligence presented as a text interface ($M = 4.16$, $SD = 1.13$). Conversely, the conditions with matching factor levels were both seen as emotionally warmer (*simple algorithms* agent as *text interface*: $M = 4.67$, $SD = 0.93$; *artificial intelligence* agent as a *human rendering*: $M = 4.83$, $SD = 1.02$). In summary, we therefore give a mixed answer to hypothesis H3. Our complex operationalization of aversion yielded some results that support the notion of artificial intelligences in human-like bodies being highly aversive—after all, the highest eeriness scores were observed for this factor combination—but participants also perceived more emotional warmth in agents with this specific design.

On an exploratory note, we would also like to report that participants' general concern about autonomous technology, which was included as a trait-like covariate, exerted a quite strong multivariate effect on the included aversion variables in our MANCOVA, $F(4,125) = 3.87$, $p = 0.01$, Wilk's $\Lambda = 0.89$, $\eta_p^2 = .11$. Further exploring this effect by means of multiple linear regression analyses, we examined that the more participants rated autonomous technology as potentially dangerous, the lower ($\beta = -.26$) was their interest to engage in future interactions, $F(1,132) = 9.15$, $p < .01$, $R^2 = .07$. Similarly, participants' ratings of emotional warmth were negatively predicted ($\beta = -.19$) by their technology concerns, $F(1,132) = 4.80$, $p = .03$, $R^2 = .04$. On the other hand, neither eeriness nor attractiveness could be connected to the general level of concern in the according analyses.

Utility Perceptions

To investigate our additional RQ, we conducted an ANCOVA with perceived agent utility as the DV and the experimental manipulation of portrayed agent mind and portrayed

agent embodiment as between-subject factors. While the main effect for portrayed agent mind (p = .56) and the interaction effect between the two factors (p = .11) did not fall below the conventional threshold of statistical significance, the main effect for portrayed agent embodiment turned out significant, $F(1,128) = 11.76$, p = .02, η_p^2 = .04. Specifically, we found that the portrayal of the agent as a human rendering led to notably higher utility ratings (M = 4.46, SD = 1.53) than the depiction of a text interface (M = 4.04, SD = 1.64). For the proposed RQ, this means that giving a human-like embodiment to digital agents might make them seem much more useful to users—keeping in mind that our previous analyses also suggest higher eeriness ratings for this kind of artificial embodiment.

Discussion

Humans have been occupied with the idea of human-like machines for a long time. In Greek mythology, Hephaestos, the divine blacksmith, built metal statues that were able to think and to feel like humans (Homer, 8th century BC). These automata assisted him in his household—in the form of "Golden Maidens"—or served as soldiers to protect the island of Crete against pirates (Talos). Today, we have entered an era in which human-like systems begin to acquire these abilities for real. Domestic robots and chatbots are among the first innovations from the realm of artificial intelligence that gain widespread use in the service sector. According to the uncanny valley hypothesis (Mori, 1970), however, technological hybrids that share human and non-human characteristics are particularly aversive whenever the human likeness of a system is high, but the machine is still distinguishable from a human.

In this research, we linked two theoretical and empirical approaches meant to explain and predict the aversion towards human-like systems. While much of the work on the uncanny valley hypothesis examined the visual design and the aesthetics of the human-like machine (e.g., Kwak, Kim, & Choi, 2017; Rosenthal-von der Pütten & Krämer, 2014), a new string of research has shifted its attention to complex digital minds (Gray & Wegner, 2012; Stein & Ohler, 2017), proposing that abilities that are deemed uniquely human, such as

emotional experience, are the main source of uncanny feelings. Connecting both approaches, we manipulated the portrayed embodiment and mind of a conversational agent and compared participants' affective response to the resulting factor combinations. Our results showed that it was mostly the embodiment of the artificial entity that drove participants' experience of uncanniness. Although the presentation of different agent minds also affected participants' eeriness perceptions, it was only in interaction with the system's embodiment that this effect manifested in a significant manner. As such, we present additional experimental evidence for assumptions formulated in (but not sufficiently addressed by) previous studies (Ferrari, Paladino, & Jetten, 2016; Stein, Liebold, & Ohler, 2019), underscoring the importance of a digital agent's visual presentation, as well as the interplay between the system's body and mind.

In a similar vein, our findings suggest that a digital agent's embodiment may exert a strong influence on both perceptions of visual *and* mental human likeness. More specifically, our results indicate that users associate complex embodiments not only with physical but also with mental prowess, even if the machine in question is actually described as a very simple system relying on question–response algorithms. Ultimately, we note that in our study, the perception of mind in a machine was much less influenced by explicit facts about its behavior and competence than by participants' impression of its visual features. Regarding the perceived warmth of the machine, we identified another interaction effect, indicating that mind without embodiment, and embodiment without mind elicited less perceived warmth in our human respondents. The Hal 9000 supercomputer in Kubrick's *2001- A Space Odyssey* comes to mind—an antagonist without a body but with its own thoughts and feelings.

Regarding the perceived utility of digital agents, we further detected an intriguing pattern. In our experiment, the agent in the human rendering condition yielded higher utility ratings, despite the fact that it was also rated as eerier by our participants. As such, it appears that particularly human-like systems increased uncanniness does not necessarily take away

from evaluations of usefulness. In our interpretation, this disconnect could be due to two

cognitive systems at work: Whereas the human-like machine elicits a fast affective response

in one system (associative system, Gawronski & Bodenhausen, 2011; Type 1, Kahnemann,

2011), slower and reflective evaluations in the second system might lead to higher perceived

usefulness (propositional system, Gawronski & Bodenhausen, 2011; Type 2, Kahnemann,

2011)—suggesting the motto "Wow, it's eerie…but actually quite useful!". Future research is

encouraged to delineate the mechanisms underlying this association, as well as its boundary

conditions.

Over and above the effects of our experimental manipulations, participants' general

concern about autonomous technologies strongly predicted lower interest in future

interactions with digital agents. This general concern is likely a function of a range of

variables, including meso- and macro-level factors (Gnambs & Appel, 2019). Specifically, we

assume that this trait is to some extent influenced by fictional depictions of autonomous

technologies. In Western cultures, science fiction is often characterized by its dystopian

outlook, leading to more critical views among potential future users (Sundar et al., 2016;

Young & Carpenter, 2018). On the other hand, more optimistic depictions of autonomous

systems and artificial intelligence might connect to reduced technology concerns. For

instance, providing a stark contrast to the West's pessimistic outlook on autonomous

machines, Japanese media are often characterized by a distinctively techno-utopian

perspective, filled with positive tropes such as the portrayal of robots as "guardian deities"

(Kaplan, 2004; Sone, 2017). In accordance with this media tonality, many Japanese people

hold quite favorable views on humanoid technologies (Hornyak, 2006), considering human-

like machines as an "efficient solution to some of the country's most pressing […]

challenges" (Ishiguro, 2017, p. 256). Even though several authors rightfully caution against

overgeneralizing Japan's society into a single, robot-affine entity (MacDorman, Vasudevan,

& Ho, 2009; Sone, 2017), it still has to be noted that the country remains the global epicenter

of robot development—highlighting the strong impact of cultural media landscapes on real-life technology acceptance. Although still sparsely, research has started to uncover similar effects with European participants as well, connecting positive technology portrayals to increased purchase intentions (Appel et al., 2016).

Limitations and Future Work

Despite the insightful contributions of our work, several limitations need to be highlighted. First, our results are based on a convenience sample and the majority of participants were female. Although gender did not affect our dependent variables directly, we found that men reported a slightly higher expertise with digital agents and fewer concerns about autonomous technologies—connections to our covariates that might, in turn, have affected our outcome criteria. As such, future research is encouraged to replicate and extend the results with a more balanced sample in terms of gender distribution.

A second limitation results from the fact that only one photorealistic human rendering was used in order to manipulate the factor agent embodiment in the current study. While the digital agent Zoe arguably constitutes a prototypical representation of the newest conversational agent technology, previous literature (e.g., Rosenthal-von der Pütten & Krämer, 2014) shows that numerous aspects of a digital agent's appearance (e.g., attractiveness, facial features, hair color) may translate into quite different user reactions, similar to the effects known from interhuman communication. This also concerns the identity of our embodied agent as a woman, which might trigger different expectations, stereotypes, and behaviors than a male counterpart. Therefore, we want to acknowledge that our study design only incorporated a small cutout from the potential spectrum of agent designs—and that our results should not be generalized to every existing embodied agent. In future studies, it might be particularly interesting to examine agent designs that appear less photorealistic or anthropomorphic than the one used in our research—such as cutely designed cartoon characters or even virtual animals.

Third, it has to be highlighted that the variables bundled together under the umbrella term *aversion* did not consistently show the expected patterns of covariation. Eeriness and ascribed emotional warmth, for example, were positively associated overall. This signals that associations regularly observed in person perception may not entirely translate to the perception of embodied agents. We believe that the dimensionality of responses towards digital agents and artificial intelligence more generally is one of the key questions to be addressed by ensuing research.

Last but not least, we would like to point out that passively observing an agent (the paradigm used in our experiment) may be subject to quite different effects than actual interactions with an agent. As such, we urge our peers to follow up on our work by including actual communication sequences between participants and agents in future studies.

Conclusion

Faced with the relentless advancement of digital technologies, people may reach quite different standpoints on how they want to treat (and be treated by) machines. As the mechanical thought experiments of previous centuries come to actual life, so do the dystopian and utopian ideas known from popular culture. In our study juxtaposing the uncanny valley of appearances with the uncanny valley of mind, we found intriguing evidence that both central aspects of the digital human—its body and its mind—clearly work together to create impressions of likable and useful, but sometimes also eerie machines. The revealed interaction effects suggest that an isolated look into the visual designs or the mental abilities of non-human entities might be a sub-optimal approach to understanding people's affinity for modern-day humanoid technologies. At the same time, we note that the uncanny experience of getting goosebumps from an advanced digital creation did not necessarily mean disliking it, which indicates a certain curiosity for the future at hand.

Acknowledgments

Declaration of interest

Declarations of interest: none.

References

Altemeyer, B. (2009). Perspectives: the decline of organized religion in Western civilization.

The International Journal for the Psychology of Religion, 14, 77–89.

https://doi.org/10.1207/s15327582ijpr1402_1

Appel, M., Izydorczyk, D., Weber, S., Mara, M., & Lischetzke, T. (2020). The uncanny of

mind in a machine: Humanoid robots as tools, agents, and experiencers. *Computers in

Human Behavior, 102*, 274-286. https://doi.org/10.1016/j.chb.2019.07.031

Appel, M., Krause, S., Gleich, U., & Mara, M. (2016). Meaning through fiction: Science

fiction and innovative technologies. *Psychology of Aesthetics, Creativity, and the Arts,

10*, 472–480. http://dx.doi.org/10.1037/aca0000052

Århem, K. (2016). Southeast Asian animism in context. In K. Århem & G. Sprenger (Eds.),

Animism in Southeast Asia (pp. 3–30). Abingdon: Routledge.

Bartneck, C., Verbunt, M., Mubin, O., & Al Mahmud, A. (2007). To kill a mockingbird robot.

In *Proceedings of the 2007 ACM/IEEE International Conference on Human-Robot

Interaction*, Arlington, USA (pp. 81–87). https://doi.org/10.1145/1228716.1228728

Broadbent, E., Kumar, V., Li, X., Sollers, J., Stafford, R. Q., MacDonald, B. A., & Wegner,

D. M. (2013). Robots with display screens: A robot with a more humanlike face

display is perceived to have more mind and a better personality. *Plos ONE, 8*(7),

e72589. https://doi.org/10.1371/journal.pone.0072589

Carpenter, R. (2018). *Cleverbot*. Retrieved from https://www.cleverbot.com

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018). In the shades of the

uncanny valley: An experimental study of human–chatbot interaction. *Future

Generation Computer Systems, 92*, 539–548. doi:10.1016/j.future.2018.01.055

Corti, K., & Gillespie, A. (2015). A truly human interface: interacting face-to-face with

someone whose words are determined by a computer program. *Frontiers in

Psychology*. https://doi.org/10.3389/fpsyg.2015.00634

Creed, C., Beale, R., & Cowan, B. (2014). The impact of an embodied agent's emotional

    expressions over multiple interactions. *Interacting with Computers, 27*(2), 172–188.

    https://doi.org/10.1093/iwc/iwt064

Dinello, D. (2005). *Technophobia! Science fiction visions of posthuman technology*. Austin:

    University of Texas Press.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical

    power analysis program for the social, behavioral, and biomedical sciences. *Behavior

    Research Methods, 39*, 175–191.

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human-machine distinctions:

    Anthropomorphic appearance in social robots as a threat to human distinctiveness.

    International Journal of Social Robotics, 8(2), 287–302.

    https://doi.org/10.1007/s12369-016-0338-y

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London, UK:

    SAGE.Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional

    evaluation model: Theory, evidence, and open questions. In J. M. Olson & M. P.

    Zanna (Eds.), *Advances in experimental social psychology* (pp. 59–127). San Diego:

    Academic Press. http://dx.doi.org/10.1016/B978-0-12-385522-0.00002-0

Gee, F. C., Browne, W. N., & Kawamura, K. (2005). Uncanny valley revisited. In

    *Proceedings of the 14th IEEE International Workshop on Robot and Human

    Interactive Communication*, Nashville, USA (pp. 151–157).

    https://doi.org/10.1109/ROMAN.2005.1513772

Gnambs, T., & Appel, M. (2019). Are robots becoming unpopular? Changes in attitudes

    towards autonomous robotic systems in Europe. *Computers in Human Behavior, 93,*

    53–61. https://doi.org/10.1016/j.chb.2018.11.045

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and

the uncanny valley. *Cognition*, *125*, 125–130.

https://doi.org/10.1016/j.cognition.2012.06.007

Griffin, M. (2019, February 10). Create digital humans in minutes, Soul Machines unveils

Digital DNA platform. *Fanatical Futurist.* Retrieved from

https://www.fanaticalfuturist.com/2019/02/create-digital-humans-in-minutes-soul-

machines-unveils-digital-dna-platform/

Guerini, M., Falcone, S., & Magnini, B. (2018). A methodology for evaluating interaction

strategies of task-oriented conversational agents. In *Proceedings of the 2018 EMNLP

Workshop SCAI*, Brussels, Belgium (pp. 24–32).Hanson, D. (2006). Exploring the

aesthetic range for humanoid robots. In *Proceedings of the ICCS/CogSci-2006 Long

Symposium: Toward Social Mechanisms of Android Science*, Vancouver, Canada (pp.

16–20).

Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. Refinement to

indices for perceived humanness, attractiveness, and eeriness. *International Journal of

Social Robotics, 9*(1), 129–139. https://doi.org/ 10.1007/s12369-016-0380-9.

Hornyak, T. N. (2006). *Loving the machine: The art and science of Japanese robots*. New

York: Kodansha International.

Ishiguro, N. (2017). Care robots in Japanese elderly care. Cultural values in focus. In K.

Christensen & D. Pilling (Eds.), *The Routledge handbook of social care work around

the world* (pp. 256–269). London: Routledge.

Jetten, J., Summerville, N., Hornsey, M. J., & Mewse, A. J. (2005). When differences matter:

Intergroup distinctiveness and the evaluation of impostors. *European Journal of Social

Psychology, 35*(5), 609–620. https://doi.org/10.1002/ejsp.282

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kamide, H., Mae, Y., Kawabe, K., Shigemi, S., & Arai, T. (2012). A psychological scale for general impressions of humanoids. Paper presented at the *2012 IEEE International Conference on Robotics and Automation,* Saint Paul, USA (pp. 4030–4037). https://doi.org/10.1109/ICRA.2012.6224790

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics, 1*(3), 465–480. https://doi.org/10.1142/S0219843604000289

Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, Chicago, USA (pp. 205-214). https://doi.org/10.1145/3171221.3171266

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PlosONE*. https://doi.org/10.1371/journal.pone.0002597

Kwak, S. S., Kim, J. S., & Choi, J. J. (2017). The effects of organism- versus object-based robot design approaches on the consumer acceptance of domestic robots. *International Journal of Social Robotics, 9*(3), 359–377. https://doi.org/10.1007/s12369-016-0388-1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Lee, K. M., Peng, W., Jin, S.-A., & Yan, C. (2006). Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication, 56*(4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

Leyens, J., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology, 31*(4), 395–411. https://doi.org/10.1002/ejsp.50

MacDorman, K. F., & Entezari, S. (2015). Individual differences predict sensitivity to the

uncanny valley. *Interaction Studies, 16*(2), 141–172.

https://doi.org/10.1075/is.16.2.01mac

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot

mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23*(4),

485–510. https://doi.org/10.1007/s00146-008-0181-2

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33–35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field].

IEEE Robotics & Automation Magazine, 19, 98–100. https://doi.org/

10.1109/MRA.2012.2192811

Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers.

Journal of Social Issues, 56(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Human Factors in

Computing Systems, 94,* 72–78.

Pagel, M. (2012). Evolution: Adapted to culture. *Nature*, 482, 297–299.

https://doi.org/10.1038/482297a

Powers, A., & Kiesler, S. (2006). The advisor robot: tracing people's mental model from a

robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART

Conference on Human-Robot Interaction*, Salt Lake City, USA (pp. 218–225).

https://doi.org/10.1145/1121241.1121280

Potter, W. J. (2013). *Media effects*. Santa Barbara: SAGE Publications.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers,

television, and new media like real people and places*. Stanford: CSLI Publications.

Reysen, S. (2005). Construction of a new scale: The Reysen Likability Scale. *Social Behavior

and Personality, 33*(2), 201–208. https://doi.org/10.2224/sbp.2005.33.2.201

Robertson, J. (2018). *Robo Sapiens Japanicus: Robots, gender, family, and the Japanese nation.* Oakland: University of California Press.

Rosenthal-von der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior, 36,* 422–439. https://doi.org/10.1016/j.chb.2014.03.066

Schultz, P. W., Zelezny, L., & Dalrymple, N. J. (2000). A multinational perspective on the relation between Judeo-Christian religious beliefs and attitudes of environmental concern. *Environment and Behavior, 32*(4), 576–591. https://doi.org/10.1177/00139160021972676

Sone, Y. (2017). *Japanese robot culture: Performance, imagination, and modernity*. New York: Springer Nature.

Stein, J.-P., & Ohler, P. (2017). Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition, 160*, 43–50. https://doi.org/10.1016/j.cognition.2016.12.010

Stein, J.-P., Liebold, B., & Ohler, P. (2019). Stay back, clever thing! Linking situational control and human uniqueness concerns to the aversion against autonomous technology. *Computers in Human Behavior, 95*, 73–82. https://doi.org/10.1016/j.chb.2019.01.021

Sundar, S. S., Waddell, T. F., & Jung, E. H. (2016). *The Hollywood robot syndrome: Media effects on older adults' attitudes toward robots and adoption intentions*. Paper presented at the 11th Annual ACM/IEEE International Conference on HumanRobot Interaction, Christchurch, New Zealand.

Tinwell, A., & Grimshaw, M. (2009). Bridging the uncanny: An impossible traverse? In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era*, Tampere, Finland (pp. 66–73). https://doi.org/10.1145/1621841.1621855

Vlahos, J. (2019). *Talk to me: Amazon, Google, Apple and the race for voice-controlled AI.* New York City, USA: Random House.

Vogeley, K., & Bente, G. (2010). "Artificial humans": Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks*, *23*, 1077–1090.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14,* 383–388. https://doi.org/10.1016/j.tics.2010.05.006

Waytz, A. & Norton, M. I. (2014). Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking – not feeling – jobs. *Emotion, 14,* 434–444. https://doi.org/10.1037/a0036054

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction, 5*(2), 29–47. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran

Young, K. L., & Carpenter, C. (2018). Does science fiction affect political fact? Yes and no: A survey experiment on "Killer Robots". *International Studies Quarterly, 62*(3), 562–576. https://doi.org/10.1093/isq/sqy028

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies, 100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Table 1. *Descriptive statistics for all dependent variables and control variables.*

Variables	Full sample	Embodiment: Text interface **Mind:** Simple algorithms	Embodiment: Text interface **Mind:** Complex AI	Embodiment: Human rendering **Mind:** Simple algorithms	Embodiment: Human rendering **Mind:** Complex AI
	$N = 134$	$n = 35$	$n = 35$	$n = 26$	$n = 38$
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Perceived mental human likeness[1]	3.74 (1.37)	3.64 (1.42)	3.37 (1.30)	3.71 (1.25)	4.18 (1.38)
Perceived visual human likeness[1]	4.31 (1.64)	3.54 (1.41)	3.10 (1.36)	5.23 (1.16)	5.51 (1.11)
Eeriness[1]	3.97 (1.02)	3.19 (0.84)	3.99 (1.08)	4.36 (0.73)	4.41 (0.89)
Attractiveness[1]	4.89 (0.91)	4.74 (0.89)	4.37 (0.74)	5.29 (0.80)	5.24 (0.91)
Emotional warmth[1]	4.53 (1.04)	4.67 (0.93)	4.16 (1.13)	4.40 (1.00)	4.83 (1.02)
Interest in future interactions[1]	3.61 (1.46)	3.60 (1.52)	3.36 (1.44)	3.71 (1.28)	3.79 (1.54)
Utility[1]	4.24 (1.59)	3.97 (1.57)	4.11 (1.72)	4.60 (1.48)	4.37 (1.57)
Expertise with digital agents[2]	2.13 (1.10)	2.07 (0.95)	2.41 (1.19)	1.96 (1.13)	2.03 (1.12)
Concerns about autonomous technology[1]	4.38 (1.06)	4.31 (0.99)	4.31 (1.01)	4.40 (1.19)	4.47 (1.12)

Notes. [1]Items were measured with 7-Point Likert scales. [2]Items were measured with 5-Point Likert scales.

Table 2. *Zero-order correlations between all obtained variables.*

		1	2	3	4	5	6	7	8	9	10	11
1	Age	–										
2	Gender[1]	.20*	–									
3	Perceived visual human likeness	.03	−.16	–								
4	Perceived mental human likeness	.07	.03	.60**	–							
5	Eeriness	.03	.06	.42**	.34**	–						
6	Attractiveness	−.14	−.15	.59**	.38**	.26**	–					
7	Emotional warmth	.01	−.12	.41**	.58**	.19*	.33**	–				
8	Interest in future interactions	.01	−.12	.39**	.57*	.16	.29**	.54**	–			
9	Utility	−.05	.10	.23**	.30**	.19*	.27**	.34**	.44**	–		
10	Expertise with digital agents	.10	.46**	−.16	.08	.02	−.03	−.01	.06	.40**	–	
11	Concerns about autonomous technology	−.22*	−.21*	.00	−.10	.10	−.13	−.19*	−.26**	−.26**	−.35**	–

Notes. * $p < .05$, ** $p < .01$. [1] Due to our coding of the gender variable, positive correlation coefficients express higher values among male participants.

Table 3. *Summary of the conducted group comparisons according to our hypotheses.*

	Main effect: Portrayed agent embodiment	Main effect: Portrayed agent mind	Interaction effect: Agent embodiment × agent mind
multivariate effect: human likeness	$\eta_p^2 = .46$***	$\eta_p^2 < .01$	$\eta_p^2 = .02$
univariate effect: visual human likeness	$\eta_p^2 = .39$***	—	—
univariate effect: mental human likeness	$\eta_p^2 = .03$*	—	—
multivariate effect: aversion	$\eta_p^2 = .25$***	$\eta_p^2 = .08$*	$\eta_p^2 = .12$**
univariate effect: eeriness	$\eta_p^2 = .17$***	$\eta_p^2 = .05$*	$\eta_p^2 = .04$*
univariate effect: attractiveness	$\eta_p^2 = .16$***	$\eta_p^2 = .02$	$\eta_p^2 = .01$
univariate effect: emotional warmth	$\eta_p^2 = .01$	$\eta_p^2 < .01$	$\eta_p^2 = .05$**
univariate effect: interest in future interactions	$\eta_p^2 = .01$	$\eta_p^2 < .01$	$\eta_p^2 < .01$
univariate effect: utility	$\eta_p^2 = .04$*	$\eta_p^2 < .01$	$\eta_p^2 < .01$

Notes. * $p < .05$, ** $p < .01$, *** p $< .001$. In the case of a non-significant multivariate effect, any further univariate testing was suspended.

Portrayed agent embodiment *(via video clips)*

		Text interface	Human rendering
Portrayed agent mind *(via vignette texts)*	**Simple algorithms**	A *web-based text interface* without any advanced emotional or cognitive abilities; all functions are based on *pre-programmed algorithms*	A *photorealistic human rendering* without any advanced emotional or cognitive abilities; all functions are based on *pre-programmed* algorithms
	Complex artificial intelligence	A *web-based text interface* able to understand and emulate complex emotions, form own thoughts, and create responses based on *complex AI technology*	A *photorealistic human rendering* able to understand and emulate complex emotions, form own thoughts, and create responses based on *complex AI technology*

Figure 1. Our study's two-factorial design, manipulating a digital agent's mind and embodiment
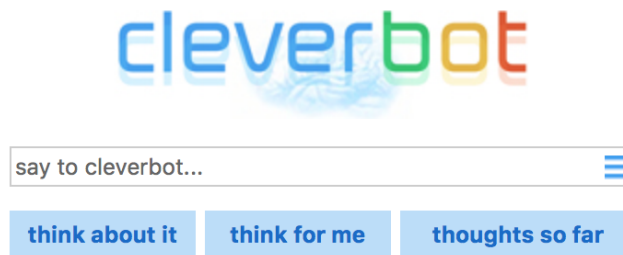

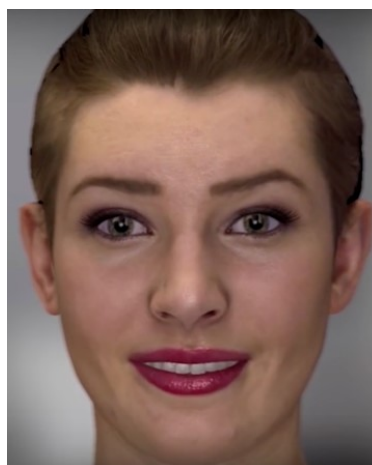*Figure 2.* Home screen of the chatbot "Cleverbot" (Carpenter, 2018) as shown in the text interface condition.

Figure 3. Digital human "Zoe" (SoulMachines, 2017) as shown in the human rendering condition.
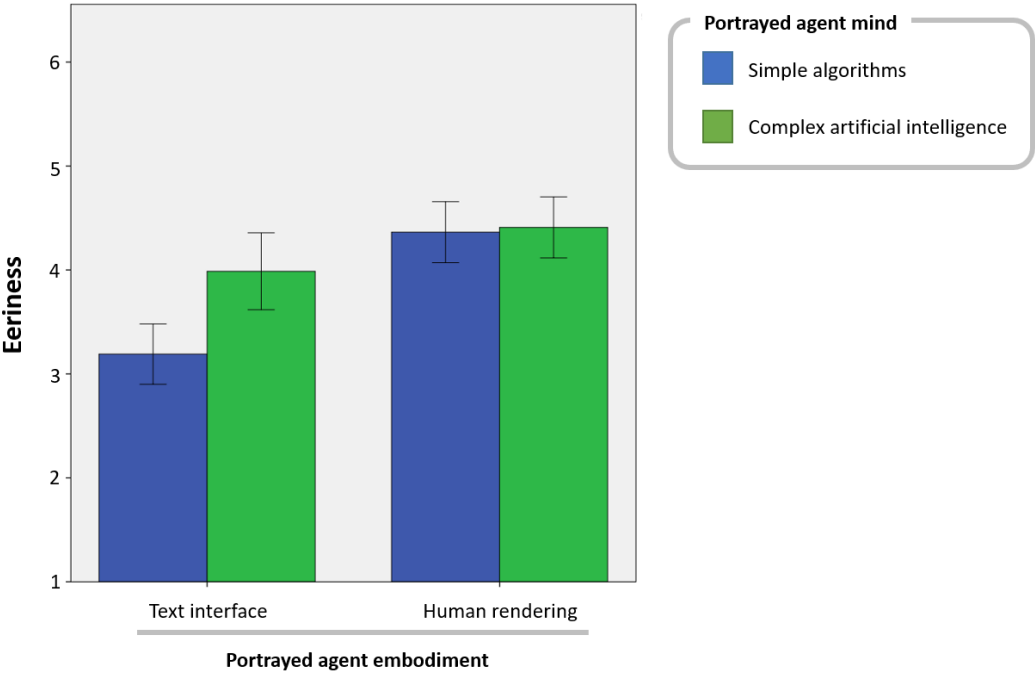
Figure 4. Interaction effect ($\eta_p^2 = .04$; *p* = .021) with the digital agent's eeriness as a DV (error bars indicate 95% CI).

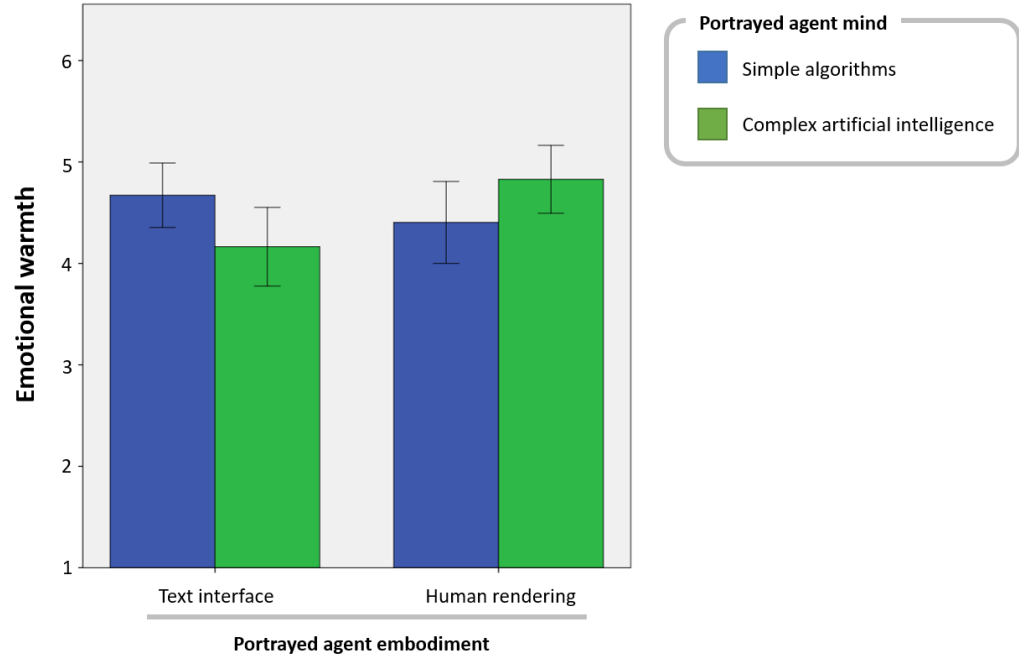*Figure 5.* Interaction effect ($\eta_p^2 = .05$; *p* = .009) with the digital agent's emotional warmth as a DV (error bars indicate 95% CI).